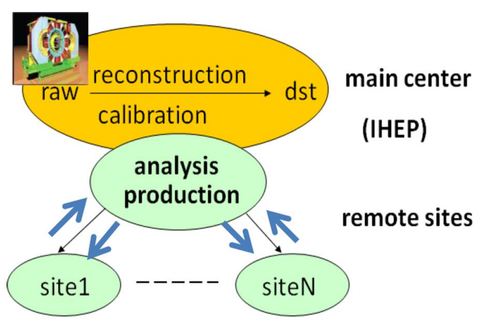
Application of StoRM+Lustre storage system in IHEP's distributed computing

Tian Yan, Xianghu Zhao, Xiaomei Zhang, Xiaofei Yan, Weidong Li

Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, P. R. China

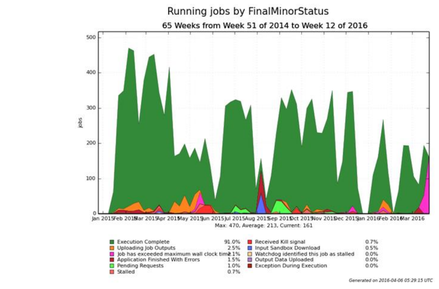
IHEP's Distributed Computing

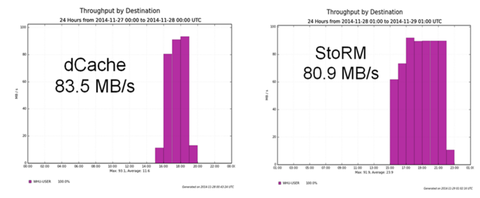
- The distributed computing system in IHEP is based on DIRAC middleware.
- It was built in 2012 originally for data production of BESIII experiment.
- In 2014, it was extended to support multi VO. Several HEP experiments such as CEPC and JUNO are joined.
- Currently, it integrates about 2000 CPU cores and 500 TB storage contributed by 16 distributed sites.
- These sites are of four types: cluster, grid, cloud and volunteer computing.



Data Production/Transfer Test

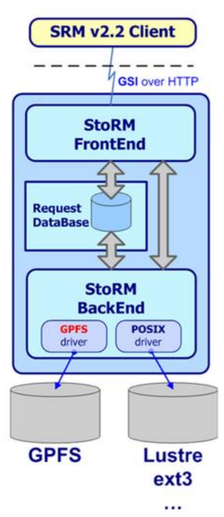
- Since Jan 2015, StoRM+Lustre has been used in CEPC data production.
- The job status figure shows that this storage works well when jobs download input data and upload output data.
- A data transfer test on WAN was performed for StoRM+Lustre, while dCache SE was chosen as a comparison.
- Test result shows that StoRM+Lustre possess the same performance as dCache in data transfer.





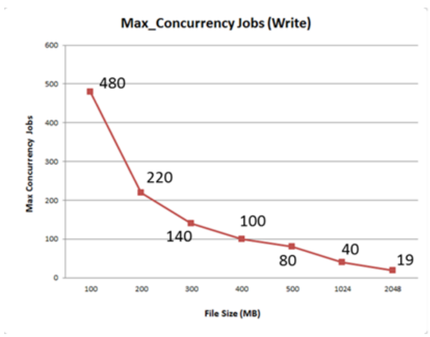
Motivation of StoRM+Lustre

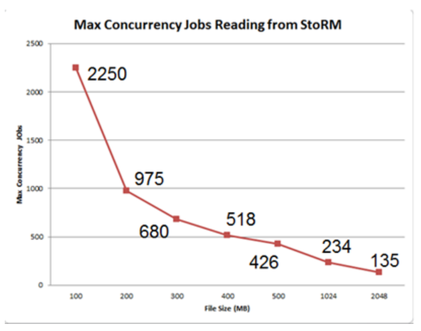
- Before StoRM+Lustre solution, we use a dCache SE with 128 TB disk array as central grid storage since 2012. While local Lustre hosted about 4PB physics data and user data.
- Data exchange between Lustre and dCache are performed manually.
- We wish to integrate local and grid storage and make the data production more automatically.
- StoRM SE was chosen because it works with POSIX file systems such as Lustre, support both grid and local access to data.
- It is lightweight, scalable, flexible and easy to deploy and maintain.

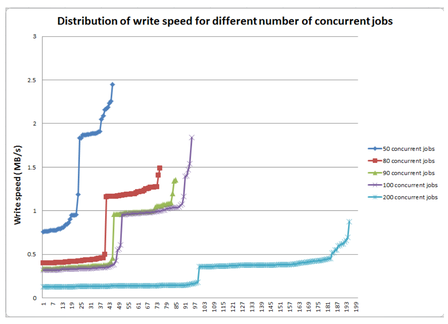


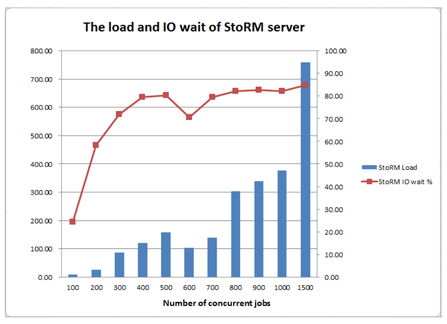
Concurrent Read/Write Test

- By using a server running many srm client threads concurrently in 10 Gbps LAN, we simulated concurrent job read/write test on StoRM+Lustre storage.
- Test result shows that the number of max concurrency jobs decrease as the tested file size increases, both for write and read.
- A "step" is observed in the distribution of write speed for each job.
- The IO resources is the bottleneck of concurrent read.
- The top priority to improve the performance of StoRM+Lustre is to increase the number of backend servers of StoRM.



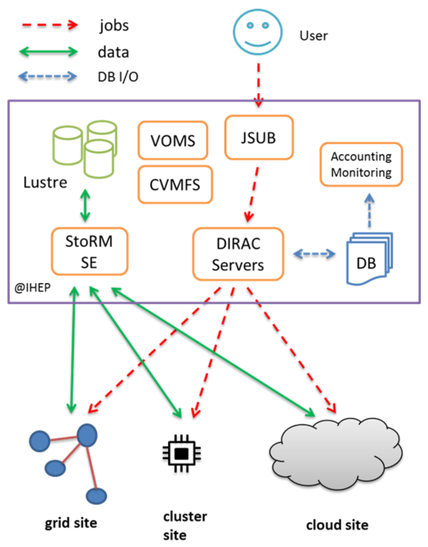






Deployment of StoRM+Lustre

- As show in the first figure, StoRM plays as a role of frontend to the grid environment, while Lustre as a backend of local accessible massive and high-performance storage.
- Users feel a nearly unified storage interface.
- Both local and remote users/jobs exchange data with Lustre storage essentially, without manually data movement between a grid SE and local storage system.
- Moreover, this architecture can be used to expose physics data in local storage to remote sites.



Summary

- In this poster, we introduced the architecture of IHEP distributed computing, explained the motivation and deployment of StoRM+Lustre storage system.
- We performed several tests, which proved that StoRM+Lustre is suitable for current data production tasks.
- This storage system is in production since Jan. 2015. And it's shown that it runs stable and plays an importance role in IHEP's data production and data exchange between sites.
- If any questions and suggestions, please contact Tian Yan (yant@ihep.ac.cn).

中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences